

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank) 2. REPORT DATE
April 7, 1995 3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Probability and Statistics Applied to the Theory of Algorithms 5. FUNDING NUMBERS
DAAL03-91-G-0110

6. AUTHOR(S)
J. Michael Steele

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pennsylvania - The Wharton School
Department of Statistics
3000 Steinberg Hall - Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104-6302 8. PERFORMING ORGANIZATION
REPORT NUMBER
JUL 17 1995

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211 10. SPONSORING/MONITORING
AGENCY REPORT NUMBER
ARO 28982.15-MA

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited. 12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
The work reviewed here is centered around the tools and results of probability theory that contribute to the theory of algorithms, especially the theory of combinatorial optimization. The reviewed cites sixteen articles and three Ph.d. theses that were supported in part by ARO Grant DAAL03-91-G-0110 during the period of May 1991 to October 1993, including the optional third year (so the present report necessarily overlaps with the report filed in October 1992 that covered the first two years of the grant). The work surveyed includes (1) work on the "objective method" which led to the the resolution with David Aldous of R. Bland's conjecture that the sum of the squares of the edges of the minimal spanning tree of a random sample from the unit square converges to a constant as the sample size goes to infinity, (2) work with R. Stine on symbolic stochastic calculus, (3) the NAS Panel Report *Probability and Algorithms*, (4) work with T. L. Snyder on the worst case behavior of the traveling salesman problem, (5) work with Jun Gao on the spacefilling curve heuristics, as well as other work on probability and its relation to the theory of algorithms.

14. SUBJECT TERMS
minimal spanning tree, stochastic calculus, probabalistic analysis of algorithms, spacefilling curves, worst case analysis, equidistribution. 15. NUMBER OF PAGES
6 16. PRICE CODE

17. SECURITY CLASSIFICATION
OF REPORT
UNCLASSIFIED 18. SECURITY CLASSIFICATION
OF THIS PAGE
UNCLASSIFIED 19. SECURITY CLASSIFICATION
OF ABSTRACT
UNCLASSIFIED 20. LIMITATION OF ABSTRACT
UL

19950703 088

Probability and Statistics
Applied to the Theory of Algorithms

FINAL REPORT

J. Michael Steele

April 7, 1995

U.S. ARMY RESEARCH OFFICE

Grant Number DAAL03-91-G-0110

University of Pennsylvania
Department of Statistics
Philadelphia, PA 19104-6302

Approved for public release; distribution unlimited. The views, opinions, and/or findings contained in this report are those of the author and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

1 Nature of the Problems Studied

The work reviewed here is centered around the tools and results of probability theory that contribute to the theory of algorithms, especially the theory of combinatorial optimization. The reviewed consists of sixteen articles and three Ph.d. theses. All of this work was supported in part by ARO Grant DAAL03-91-G-0110 during the period of May 1991 to October 1993, including the optional third year; so the present report necessarily overlaps with the report filed in October 1992 that covered the first two years of the grant.

2 Summary of Major Contributions

The Objective Method. Probably the best of the work that was initiated with the assistance of this grant is represented by the article "Asymptotics of Euclidean Minimal Spanning Trees on Random Samples," written with David Aldous and published in *Probability and Related Fields*.

This article resolved an interesting conjecture of R. Bland's that the sum of the squares of the edges of the minimal spanning tree of a random sample from the unit square converges to a constant as the sample size goes to infinity. Although it is reinforcing to have resolved Bland's conjecture, the main contribution of the article is to provide a compelling example of the "objective method" whereby one obtains limit results by constructing an "infinite analog of the object of interest." The justification that one has the right infinite analog then requires showing that the proposed candidate provides concrete insight into the large sample behavior of the original finite problem. In the problem addressed with Aldous, both parts of the program were brought to a satisfying conclusion.

This is a rich method that continues to make contributions to other problems. Aldous and I have found several additional examples of the method that show its potential and that show that the method deserves to be distinguished from the methods of weak convergence, or invariance principles. Naturally, there are links between the theory of weak convergence and the "objective method," but there are also distinctions that merit being preserved. Among these distinctions, the most important is probably that the objective method usually requires that one *invent* some infinite object, while in the theory of weak convergence one focuses more on the finite objects and then *identifies* the limit under suitable growth conditions and with respect to a suitable metric.

| | |
|--------------------|--|
| on For | |
| RA&I | <input checked="checked" type="checkbox"/> |
| S | <input type="checkbox"/> |
| need | <input type="checkbox"/> |
| ation | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Dist | Avail and/or Special |
| A-1 | |

Since the publication of the article with Aldous, there have been several important contributions to the MST problem. First, the central limit theorem conjectured in Aldous and Steele (1992) has been proved by K. Alexander (1995), and subsequent to that work, Kesten and Lee (1994) proved a CLT for the minimal spanning tree with power weighted edges in all dimensions. The method of Alexander was closely involved with an earlier method of D. Ramey and called upon the methods of percolation theory, but the apparently more powerful method of Kesten and Lee is closer to martingale methods.

Symbolic Methods and Itô Calculus. The article "Applications of *Mathematica* to Stochastic Calculus" written with R. Stine reports on an implementation of the basic formalism that arises when one implements tools for the stochastic calculus and Itô's formula in the *Mathematica* platform. The resulting tools greatly shorten many of the routine calculations of the stochastic calculus.

The first report on this software development was given in the *Proceedings of the ASA Section on Computational Statistics*, but a much expanded and more polished version "Mathematica and Diffusions" was written for *Economic and Financial Modeling with Mathematica* published in 1994 by Springer-Verlag (Hal Varian, editor). The *Mathematica* code developed in these articles is distributed on disk with the Springer-Verlag volume.

Probability and Algorithms: A Panel Report. The National Academy of Sciences has published "Probability and Algorithms: A Panel Report," and the report was also reprinted in February of 1993 as a special issue of *Statistical Science*. I served as the chairman of this panel and contributed to the report's development in project conception, panel selection, project organization, and final editing. David Aldous and I wrote the introduction to the report; and, in addition to writing a chapter that addresses problems in Euclidean Combinatorial Optimization, I wrote the concluding chapter.

Equidistribution and other Worst Case Problems. The article "A priori Bounds for the Euclidean Traveling Salesman Problem," written with T. L. Snyder continues one of my basic lines of research: the development of inequalities for heuristics for NP-hard problems. This article was presented as an invited paper at the *Eighth Symposium in Computational Geometry* in Berlin in June 1992. The article was subsequently extended and refined

for publication and the article is scheduled to appear in *SIAM J. Discrete Math.*.

Spacefilling Curve Heuristic. The spacefilling curve heuristic is a method for studying problems of combinatorial optimization for points in the plane where one begins by transforming the planar points into points on the line by means of an inverse to continuous spacefilling curve. The method was explored extensively in the Princeton University thesis of Jun Gao that was supported by this grant, and two articles that engage the spacefilling curve heuristic have been published with the support of this grant.

Other Results. The other contributions reviewed below are more specialized.

1. The article "Convex Hulls of Random Walks" which was also written with T. Snyder was published in the *Proceedings of the American Mathematical Society*. The main point of the article is to provide an exact formulae for the length of the convex hull of $\{S_1, S_2, \dots, S_n\}$ where $S_k = X_1 + X_2 + \dots + X_k$ and the X_i are i.i.d. with values in \mathbb{R}^2 .

The results are surprisingly simple and they offer an engaging instance of combinatorial and geometric methods coming together in succinct and definitive results. The work was motivated by earlier contributions by Spitzer and Baxter.

2. The article "Transient Behavior of Coverage Processes" written with S. Browne responded to a target of opportunity. The problem that we solved is that of determining the distribution of the clump size in a coverage process where the first interval in any clump has a distribution that differs from that of the subsequent intervals. The distributional question comes up in a variety of contexts including storage problems, scheduling problems, and queuing problems where the first service of any busy period has a exceptional service length distribution. The article has been accepted by the *Journal of Applied Probability*.
3. The article "Probabilistic Networks and Network Algorithms" with T.L. Snyder is scheduled to appear as part of the North-Holland Series *Handbook of Operations Research*. The article has both expository and research components. In particular it aims to provide a basic

grounding suitable for a researcher who would wish to obtain a state-of-the-art view of probabilistic network theory.

3 Research Articles Supported by this Grant

1. "Asymptotics of Euclidean Minimal Spanning Trees on Random Samples," (with David Aldous), *Probability Theory and Related Fields*, **92**, (1992), 247-258.
2. "Euclidean Semi-matching of Random Samples", *Mathematical Programming*, **53**, (1992), 127-146.
3. "A Priori Bounds for the Euclidean Traveling Salesman Problem," (with T. L. Snyder), *Proceedings of the Eight Annual Symposium: Computational Geometry*, ACM Press, New York, NY, 1992, 344-349.
4. "Introduction to Probabilistic Algorithms and the Probabilistic Analysis of Algorithms," (with David Aldous), *Probability and Algorithms: A Panel Report*, 1-15, National Academy of Sciences, National Research Council, Board on Mathematical Sciences, Washington D.C. 1992.
5. "Probability and Problems in Combinatorial Optimization," *Probability and Algorithms: A Panel Report*, 109-129, National Academy of Sciences, National Research Council, Board on Mathematical Sciences, Washington D.C. 1992.
6. "Missing Pieces, Derandomization, and Concluding Remarks," *Probability and Algorithms: A Panel Report*, 175-178, National Academy of Sciences, National Research Council, Board on Mathematical Sciences, Washington D.C. 1992.
7. "Applications of Mathematica to Stochastic Calculus," (with R. A. Stine), *ASA 1991 Proceedings of the Statistical Computing Section*, (1991), 11-19.
8. "Transient Behavior of Coverage Processes with Applications to Infinite Server Queues," (with S. Browne), *Journal of Applied Probability*, **30**, (1993), 589-601.

9. "Convex Hulls of Random Walks," (with T. L. Snyder) *Proceedings of the American Mathematics Society*, 117, (1993), 314-324.
10. "Probabilistic Networks and Network Algorithms," (with T. L. Snyder). To appear in *Handbook of Operations Research*, North-Holland, Publishers.
11. "General Spacefilling Curve Heuristics and Limit Theory for the Traveling Salesman Problem," (with Jun Gao), *Journal of Complexity*, 10 (1994), 230-245.
12. "Sums of Squares of Edge Lengths and Spacefilling Curve Heuristics for the Traveling Salesman Problem," (with J. Gao), *SIAM J. Discrete Mathematics*, 7, (1994), 314-324.
13. "Mathematica and Diffusions," (with R. A. Stine) in *Economic Modeling with Mathematica*, Hal Varian, ed., 192-213, Springer-Verlag, New York, 1993.
14. "Equidistribution in all Dimensions for Worst-case Point Sets for the TSP," (with T. L. Snyder). To appear in *SIAM J. Discrete Math.*
15. "Variations on the Monotone Subsequence Problem of Erdős and Szekeres." To appear in *Discrete Probability and Algorithms*, IMA Volumes in Applied Mathematics, Vol. 72, 1995.
16. "Three Statistical Technologies with High Potential in Biological Imaging and Modeling" (with M. Fridman) in *Monte Carlo Methods in Radiation Biology*, M. Varma and A. Chatterjee (eds.), Plenum Press, New York, 1995.

4 Participating Scientific Personnel and Advance Degrees Earned

1. J. Michael Steele (Principal Investigator)
2. J. Gao. Mr. Gao was supported under this grant for two years of his graduate work at Princeton University. Mr. Gao completed his Princeton Ph.D in September, 1992, and his dissertation "Analysis of

Two Heuristic Methods for the Euclidean Traveling Salesman Problem" has been communicated to the ARO Library. Two articles related to this thesis have been discussed above.

3. Moshe Fridman. Mr. Fridman completed his University of Pennsylvania Ph.d. in 1993 with a thesis "Hidden Markov Model Regression" which has been transmitted to the ARO Library.
4. Nanda Piersma. Ms. Piersma completed her University of Tinbergen Doctoral degree in 1993. She received partial support from this grant when she visited the University of Pennsylvania for two prior summers and part of the 1991-92 academic year.
5. Claude Athaide. Mr. Athaide began his thesis work under my direction, and during my leave last year at the IMA at the University of Minnesota Mr. Athaide completed his work with Robert Stine. His thesis "Likelihood Evaluation and State Space Estimation for Non-linear State Space Models" has been communicated to the ARO Library.

5 Additional References

Alexander, K. (1994) "The RSW Theorem for continuum percolation and the CLT for Euclidean minimal spanning trees", Technical Report.

Kesten, H. and Lee S. (1995) "The Central Limit Theorem for Weighted Minimal Spanning Trees on Random Points," Technical Report.